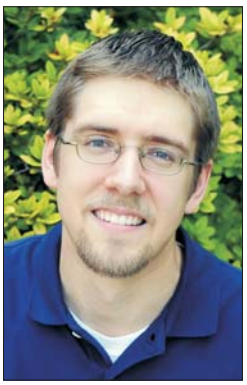
Biological Information, Molecular Structure, and the Origins Debate

Jonathan K. Watts



Jonathan K. Watts

Biomolecules contain tremendous amounts of information; this information is "written" and "read" through their chemical structures and functions. A change in the information of a biomolecule is a change in the physical properties of that molecule—a change in the molecule itself. It is impossible to separate the information contained in biomolecules from their structure and function. For molecules such as DNA and RNA, new information can be incorporated into the sequence of the molecules when that new sequence has favorable structural and functional properties. New biological information can arise by natural processes, mediated by the interactions between biomolecules and their environment, using the inherent relationship between structure and information. This fact has important implications for the generation of new biological information and thus the question of origins.

A traveler is checking in for a flight and her bags are slightly over the weight limit. Without hesitating, she pulls out her iPod. It is very heavy, she explains to the check-in agent, since it contains thousands of songs. She deletes most of the music, repacks the iPod, and reweighs the bags—which are now well within the weight limits.

Or consider a kindergarten student, learning to write letters. He writes a whole page of A's with no trouble. Next he wants to practice writing the letter G. But after a few G's are written, they seem to want to fold onto each other, as though he were writing on the sticky side of a piece of tape. Each new G he manages to add contributes a new wrinkle or fold, until eventually he gives up and decides to practice writing a less troublesome letter.

When we laugh at these two impossible stories, it reveals how deeply, almost reflexively, we tend to feel that information should be distinct from physical properties. At least in terms of computer code or printed text, we expect that similar devices containing different information will have similar physical properties. By contrast, different devices may contain the same information in spite of their dramatically different physical properties (for example, the printed and online versions of this article).

But biological information is quite different. This article will show that there is a fundamental difference between biological information and abstract information such as computer code or text: the biological information cannot be separated from its structure. The structure and reactivity of biomolecules can give rise to new informa-

Jonathan Watts *is a postdoctoral fellow in the Departments of Pharmacology and Biochemistry at the University of Texas Southwestern Medical Center at Dallas. He received his BSc in chemistry from Dalhousie University (Halifax, Canada) and then spent a year on mission in Côte d'Ivoire before returning to science and obtaining his PhD in nucleic acid chemistry at McGill University (Montreal, Canada) in 2008. His research interests center on the chemistry, chemical biology, and therapeutic applications of oligonucleotides. He has published eighteen papers. Jon and his wife Katie have two preschool children, Liam and Lydia. He enjoys music and often leads worship at his church. Correspondence can be sent to jonathan.watts@ymail.com.*

tion without the direct input of an intelligent agent. Thus we need to be careful when analogies from the world of computers or literature are applied to biological information. This is important in terms of the debate on the origin of life.

The Information-Structure Duality of Biomolecules

Discussion of biological information is often limited to the DNA (or RNA/protein) sequence, which superficially looks much like the kinds of abstract information we are familiar with. When the human genome sequence was published, biology was said to have entered an information age. Stephen Meyer begins his book *Signature in the Cell* by quoting from sources as diverse as Bill Gates and Richard Dawkins who find that "the machine code of the genes is uncannily computer-like." Meyer's next question is highly pertinent: "If this is true, how did the information in DNA arise?"[1]

While I enjoyed reading much of *Signature in the Cell,* I felt that the analogy between DNA and abstract information was taken too far. The issue is that biological information is not abstract: it is always mediated and interpreted by physical interactions. While studying the chemistry and biochemistry of oligonucleotides (short sequences of DNA, RNA, and their chemically synthesized analogues), I have often come face-to-face with the frustration that can be caused by forgetting how tightly information and structure are intertwined.

Some oligonucleotide sequences can be manipulated easily enough, such as the letters within an abstract line of text. But other sequences have repeatedly reminded me that a DNA sequence is not just an abstract line of text. For example, a famous sequence called the Dickerson-Drew dodecamer (5'-CGCGAATTCGCG)[2] can bind another copy of itself by classic Watson-Crick base pairing (A-T and G-C pairs, figure 1a). But under different conditions, it will instead fold back on itself, forming into a "hairpin" structure while still making use of Watson-Crick base pairs (figure 1b). Various factors, including chemical modifications, can favor one structure over the other.[3] While the sequence information is the same, the two structures respond very differently in experiments (i.e., they exert different functions).

Some of my colleagues have made various chemically modified analogues of the sequence GGTTGGTGTGGTTGG.[4] Since this sequence contains only one half of each possible Watson-Crick base pair, one might expect that it would behave "properly" and exist as a nice unstructured line of chemical "letters." On the contrary, it folds into a very complex structure having nothing to do with Watson-Crick base pairing (figure 1c).[5]

The two stories we began with were not chosen at random. Separating and characterizing biomolecules by their mass, for example, is one of the simplest ways to analyze their information content. After synthesizing an oligonucleotide, I first analyze it by gel electrophoresis (here, my desired sequence and any impurities that may be present are separated according to their mass as they are pulled through a gel by an electric field). Then, before carrying out experiments with the oligonucleotide, I inject a small part of each sample into a mass spectrometer to determine its mass more precisely. If the mass of a synthetic oligonucleotide is correct, we can generally assume that the sequence is what we were trying to produce (i.e., the oligonucleotide contains the expected information).[6]

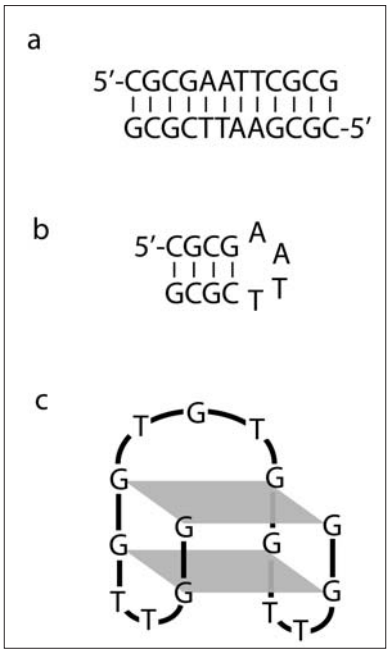


Figure 1. Some Structures of Oligonucleotides.
(a) The duplex form of the Dickerson-Drew dodecamer, which is in equilibrium with (b) the hairpin form. One form may predominate, depending on the experimental conditions and the chemistry of the sequence. Both are based on Watson-Crick base pairing: G with C and T with A.
(c) An example of a highly stable structure having nothing to do with Watson-Crick base pairing. The G residues arrange themselves in "quartets" in a stacked planar arrangement.

The second story, as you may already have guessed, relates to the DNA bases A and G, adenine and guanine. While both are purine bases and are closely related, guanine folds into a much greater variety of structures and binds to itself with high affinity (as in the sequence from figure 1c). It is nearly impossible, using standard biochemical techniques, to copy a DNA sequence containing dozens of adjacent G's. On the other hand, sequences containing dozens of A's are easy to copy and are used each day in laboratories around the world ("polyA" sequences similar to these are also added to the ends of all of the messenger RNA in our cells).

In a companion article in this issue, Randy Isaac explores the nature of biological information.[7] He points out that when there is an abstract linkage between a given type of information and its meaning, we can readily identify that the information was directly written by an intelligent agent. In contrast, when the linkage between information and its meaning is entirely physical (for example, molecular structure, or function mediated through thermodynamic interactions), we may not be able to attribute intelligent agency as quickly.

So, in thinking about biological information as it relates to the origin of life, we must be careful with analogies from the familiar world of computers or books. The information in a book can be stored in multiple physical forms: a large-print hardcover edition, an electronic PDF version, or even Braille. When we read it with the appropriate media or tools, we obtain the same information. In contrast, biological information cannot be separated from its structure. Three different representations of the nucleobase adenine are shown in figure 2. The first representation, A, is a common abbreviation used in sequence analysis. It is a letter, a symbol, of the biological information carried by adenine. This simple representation facilitates communication and information transfer among researchers. In the second representation, the various atoms are specified. Much more information is included here—the types of atoms contained in adenine and the arrangement of bonds that hold it all together. Chemists would be very comfortable with the second representation. But the readers and writers of biological information (enzymes or other nucleic acids, for example) have to work with something even more complex, something much more similar to the third structure: a three-dimensional electron surface with a defined shape and regions of positive and negative charge.

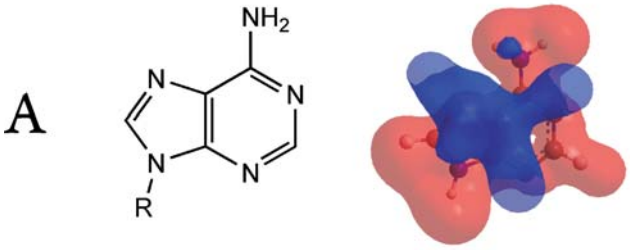


Figure 2. Three Ways of Representing the Nucleobase Adenine. Left, the letter "A," as commonly used when discussing the DNA sequence. Center, the chemical structure of adenine, showing the atoms that make up adenine, their spatial arrangement, and the types of bonds that connect them. "R" represents the sugar-phosphate backbone. In keeping with organic chemistry convention, carbon is assumed to be at any corner not labeled with a different letter, and carbon-bound hydrogens are left out. Right, a computed model of adenine, showing electron surfaces of net positive or negative charge (light gray or dark gray, respectively, in the print version of this article; red or blue, respectively, in the PDF version of this article). The model was generated using Gaussian03W and Chem3D.

Figure 2 gives three levels of understanding of the information conveyed by a single "A" in the DNA sequence. The one on the left looks something like computer code or text, but, in fact, the complex electronic structure on the right is what enzymes and other "information readers" have to interpret. There is much more information in this full structure, but it is much harder to quantify and looks nothing like text or code. Perhaps surprisingly, taking this more complex view of biological information and its connection to structure will make it easier to see and to understand how new functional information can be generated without being directly written by an intelligent agent.

The Generation of New Sequence Information from Structural or Functional Components of Biomolecules

William Dembski[8] and others in the intelligent design (ID) community[9] claim that natural causes are insufficient to produce complex specified information. Their "law of conservation of information" can, like any law, be disproved if examples are found that violate the law. Yet I find that the law as formulated by Dembski does not work in the laboratory;

information can and does arise without direct intelligent input.

Much of the response to the idea of such a law[10] has discussed the information that arises through processes of mutation and natural selection.[11] Others have written about the new information generated by the immune system when it is presented with an antigen.[12] In these cases, information (and associated function) is not directly written by an intelligent agent, but arises from the interplay of an organism with its environment.

In *Signature in the Cell,* Meyer restricts his version of the law of conservation of information to a nonbiological starting point.[13] In keeping with this context, I will also discuss a nonbiological example that is commonly encountered in both academic and corporate research labs. New information can arise from the structure of a molecule such as DNA and the molecules it interacts with.

To begin this experiment, a random oligonucleotide is made on an automated gene synthesizer. These instruments are usually used to make specific (nonrandom) sequences, which can be programmed into the instrument according to what the scientist specifies. The instrument goes through a "synthetic cycle" for each successive nucleotide in the chain—adding one nucleotide at a time, drawing from the appropriate choice of four vials: one for each of A, T, G, and C. For our experiment, we will adapt the instrument to make a random oligonucleotide sequence by simply combining the four building blocks in a single vial so that all are equally likely to be incorporated at each coupling step. Repeating the synthetic cycle, say fifteen times, would yield an oligonucleotide 15 nucleotides long. There are 4^{15} (or just over a billion) different 15-nucleotide sequences. At a typical synthesis scale (25 nanomoles), about 10 million copies of each different option would be present. So far we have lots of complexity but no specificity. In other words, there are a lot of sequences carrying a lot of information,[14] but no useful, *functional* information is present because we have not chosen between any of the options.

However, we can provide specificity by selecting sequences according to their structure or function. For example, our pool of random 15-nucleotide sequences will likely contain GGTTGGTGTGGTTGG, the oligonucleotide from figure 1c. This complex structure binds tightly to the protein blood-clotting factor thrombin. If we wash our entire pool of random oligonucleotides across a sample of thrombin, this sequence will stick more tightly than others (figure 3). This is based on a real example: the sequence GGTTGGTGTGGTTGG was not rationally designed to bind thrombin, it was discovered by a similar

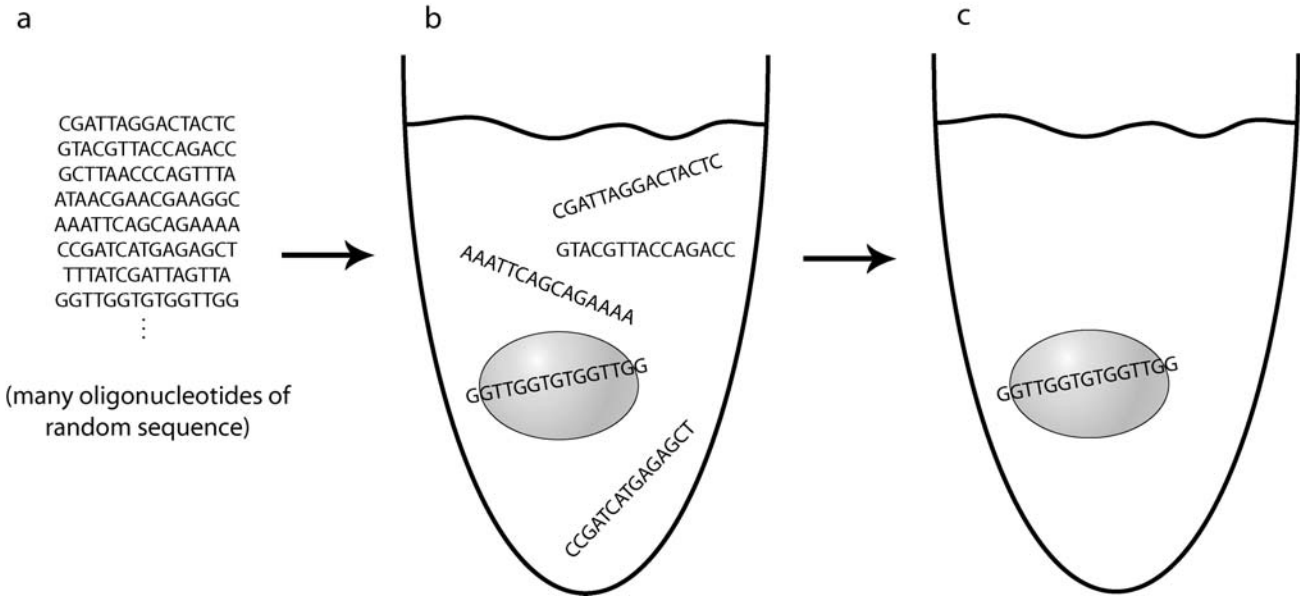


Figure 3. *In vitro* selection of an oligonucleotide consists, at its simplest, of (a) generating random sequences, then (b) selecting and identifying sequences with the desired properties from the pool. Here we show protein binding as a selection step; the sequences that do not bind are removed as shown in part (c). Those sequences that bind their target may be amplified (copied) and then undergo the selection cycle several more times. In this way, the best candidates can be identified.

experiment to the one I have just described.[15] Yet it clearly has generated or uncovered a DNA sequence with specified information.

Variations of this technique are commonly used to develop DNA, RNA, or even proteins with desired properties.[16] Beyond a simple function such as binding to a given target, it can produce more complex functions such as catalysis of a chemical reaction.[17] The general process is called *in vitro selection*, or sometimes *in vitro evolution* or *SELEX* (Systematic Evolution of Ligands by EXponential enrichment).[18] Generally, the selection cycle is repeated several times to improve the signal-to-noise ratio and to identify oligonucleotides with the very best properties. In between each repetition of the selection step, the surviving oligonucleotides are copied ("amplified"). SELEX-derived sequences have proven their usefulness as probes to bind target proteins and small molecules alike,[19] and have even led to an FDA-approved therapeutic.[20]

Objections Addressed

Meyer claims that substantial amounts of information are put in by scientists during an *in vitro* selection process, to the extent that negligible net information is really produced.[21] For example, random oligonucleotides are often synthesized with "wings" attached at either end, consisting of a known sequence. This helps the experimenter to amplify the selected sequences (copy them in sufficient quantity for further use). This amplification is typically done using information-rich enzymes. And the selection step itself is designed by the experimenter.

Let us take these objections one at a time, and I will try to explain why either they are not strictly necessary to a SELEX experiment, or they do not count as an inappropriate introduction of information. First of all, the wings of a known sequence are used for making copies of our selected sequences and for measuring the sequence information.[22] This is an analytical problem: the sequences with higher affinity have *already* been selected by their binding to the protein, and thus we already have a certain degree of new, functional, specified information, even before we amplify or read the sequences. And, of course, whether or not wings of a known sequence are present during a selection, a *region of genuinely random sequence* is being selected and yields new specified information. The constant wings are present both before and after and so do not count against the new information generated.

What about a polymerase enzyme used to make new copies of the selected sequences at each amplification step? First of all, progress is being made toward enzyme-free amplification of nucleic acids, so an enzyme may someday not be required to amplify our sequence of interest.[23] Otherwise, all of the same responses can be given at this point. *The selection of information has already taken place when one sequence binds its target with higher affinity than others;* thus the amplification and sequencing are again simply analytical tools. And finally, the information contained within the enzyme is unchanged and remains constant throughout the selection; thus its presence does not detract from the fact that new information is being obtained.

Finally, what about the information put in by designing and executing a series of selection steps? Scientists carefully design SELEX experiments, it is true. However, I think there are at least three reasons why this objection does not stand.

First, the key selection step actually occurs when one oligonucleotide binds its target to a greater extent than others. This is a purely physical process and does not depend on investigator input.

Secondly, while amounts of information can be hard to quantify when comparing different types, it is hard to argue that a short series of manipulations, moving liquid from one tube to another, contributes anything similar to the amount of information contained in, for example, a 15- to 60-nucleotide chunk of DNA of a specific and functional sequence.

Thirdly, it is not always necessary for researchers to intervene at each step, showing the parallel between SELEX and putative natural examples of molecular evolution. For example, two groups have demonstrated systems for the continuous *in vitro* evolution of biomolecules.[24] In these two different examples, in place of a series of selection steps, a system is designed so that biomolecules (RNA and proteins) are continually optimized through mutation and replication, and the best sequences are preserved. Continuous *in vitro* evolution is very closely related to natural selection. Thus we have come full circle: *in vitro* selection steps mimic natural selection,

something that clearly does not require direct human input. In the simplest SELEX experiment, oligonucleotides that confer a needed function (say, binding to a target) survive (by being copied and identified). In nature, the functions may be different, but survival and reproduction are still just as relevant. Thus the selection carried out by a researcher to obtain oligonucleotides with desired properties is parallel to the selection pressures of the environment on any adapting molecule or organism when a new generation survives and multiplies.

In summary, the interventions and manipulations by researchers have parallels in natural selection and biomolecular evolution—diversification by mutation, selection by survival, repetition. The key selection step—that is, the step that specifies information from the complex-but-random starting pool—occurs by the interaction of a biomolecule with its environment, not by intervention by the researchers. Even if researchers set up certain conditions, the desired sequence is unknown by any intelligent agent involved in the experiment. Useful, functional, specified information is generated from a random starting point.

Is the amount of information generated in a SELEX experiment so small as to be negligible? A specific 20-nucleotide sequence corresponds to 40 bits of information.[25] There are hundreds of examples of functional oligonucleotides generated by *in vitro* selection.[26] Thus *in vitro* selection experiments have generated thousands of bits of information over the past two decades.

Where Does the Information Come From?

At the end of a SELEX experiment, a biomolecule contains more information than the researchers put in. Is there another source for this information? Yes. During a SELEX experiment, *information from the environment is captured in the form of a particular DNA or RNA sequence*.[27] This information transfer works because of the relationship between structure and information.

The surface of the target contains information about the positions and charges of a huge array of electron orbitals (something similar to figure 2c, but much larger and more complex). This information is mirrored in the structure of a particular oligonucleotide that folds in a unique way, and the match allows the two molecules—DNA and target—to bind. We use the relationship between the DNA and the target protein to transfer the information into a form we can easily amplify, read, and reproduce—a DNA sequence. The same principles apply when we select an oligonucleotide that catalyzes a chemical reaction or binds to a small molecule rather than to a protein.

Molecules are constantly interacting with each other. Most of the time they "bounce off" one another, but occasionally they bind together, or even undergo a chemical reaction with each other. The interactions between molecules are information-rich (for example, as any chemist will tell you, sometimes the reactivity of a molecule can be used to identify its structure). So why have I focused this article on the transfer of information into a sequence of DNA or RNA? Nucleic acids such as these are a wonderful medium for molecular evolution because they are so easy to copy and analyze. No other type of complex molecule that we know of can be synthesized chemically, copied enzymatically, and sequenced so readily.

However, various creative researchers have nonetheless found ways to evolve other types of molecules and reactions in the laboratory. For example, one strategy is to tether reactive molecules to short pieces of DNA: when two particular groups are joined under a particular set of reaction conditions, they leave a trace in the DNA sequence that can be amplified and measured.[28] This has led to the discovery of new types of chemical reactions.[29]

SELEX, Serendipity and Complexity

SELEX experiments are so useful precisely because of their ability to capture so much information. In fact, one reason scientists incorporate randomness and evolution into our discovery efforts is that reality is often too complex for our attempts at the alternative: rational design.[30] We allow chance and selection room to work (in this case, by beginning with a random oligonucleotide). While SELEX is only twenty years old as a technique, the idea of the importance of serendipity in science is much older.

In the same way, researchers on *both sides* of the origins debate must recognize that the science of origins is too complex for our attempts to understand. Indeed, origin-of-life researchers themselves are often the first to admit that they do not understand how life originated. Just because we can generate biomolecules containing new information in a SELEX experiment does not mean we are anywhere near understanding or recapitulating the origin of life.

The ID community should also recognize the limits of our knowledge. Biological information is too dynamic to support a law of conservation of information. Hard lines cannot easily be drawn between the information in biomolecules and the information in the rest of the environment. Substantial empirical evidence shows that biological information increases through natural causes; SELEX provides one example of such an increase. When information is properly understood in its connection to biomolecular structure, it is not surprising that new biological information can arise from natural processes. Thus the structural component of biological information adds another level of complexity to the origins debate. Biological information is too complex and too dynamic for us to be able to make probabilistic claims of a "designed" origin based on the amount of information contained in biological systems today.

Meyer and Dembski claim that the probabilistic resources of the universe are simply not sufficient to allow the generation of information-rich self-replicating biomolecules.[31] However, an evaluation of the probability of a sequence arising depends almost entirely on our knowledge of the mechanisms whereby such an event may occur.[32] For example, Wilf and Ewens have shown that the probability of generating a given sequence depends strongly on whether the sequence is independent of history (as in a coin toss) or can preserve advantageous elements from "ancestor" sequences (as in many types of both SELEX and natural selection).[33]

Meyer claims that the argument for direct intelligent design of DNA is not based on an absence of knowledge, but a knowledge of absence.[34] Yet, if the ID community responds to the points I have made here, they will likely do so using gaps: "No one knows how random oligonucleotides could self-assemble to provide a starting pool on which prebiotic selection could act." "No one knows how early RNAs could replicate in the absence of polymerase enzymes." These statements are currently correct—but rapid progress is being made in both areas.[35] It would simply not be true to say, "We *know* that random-sequence oligonucleotides cannot self-assemble" or "We *know* that enzyme-free RNA replication is impossible." Thus, in spite of his objections to the contrary, Meyer's arguments about generation of biomolecular information at the origin of life are substantially based on absence of knowledge.[36]

Conclusions

We must be careful when comparing biological information to familiar forms of information such as text or computer code. Biological information is not abstract; it is intimately tied to the structure and function of biomolecules. As such, the biological information in cells can increase through natural processes. Perhaps the first cell was created out of nothing—but the high information content of modern cells does not prove this "special creation" of the first life. Another option is that processes closely or distantly analogous to SELEX could have been used to increase the amount of information in a primitive replicating system, although science has not yet identified such a system. A sense of wonder and worship of the Creator is appropriate in either case.

As a Christian I believe deeply and thoroughly in design. But that design does not oppose the fact that both organisms and molecules can accumulate information through natural processes. When I read about experiments in molecular evolution, I am often inspired by the complexity and beauty of the biomolecules that can generate new information by interacting with their environment. I am also inspired by the creativity of the researchers who did not *directly* design new sequences, but set up a system in which they could measurably evolve. I see unmistakable parallels in God's activity in the world—the beauty and complexity around us speaks of God's subtlety and majesty, even as there is abundant evidence that molecules and organisms can generate new information through physical interactions with their environment.

It is essential that we avoid the false dichotomy of "things God did" versus "things science can under-

stand." In all of our research, including questions of origins, we should worship God in both the places of our ignorance and of our knowledge. The gaps in our knowledge should lead us to greater humility and thus worship. Likewise, each new discovery opens our eyes to new depths of beauty in creation, and these should also lead us to worship.

Notes

[1]S. Meyer, *Signature in the Cell* (New York: HarperOne, 2009), chapter 1.

[2]R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, and M. L. Kopka, "The Anatomy of A-DNA, B-DNA, and Z-DNA," *Science* 216, no. 4545 (1982): 475–85.

[3]J. K. Watts, B. D. Johnston, K. Jayakanthan, A. S. Wahba, B. M. Pinto, and M. J. Damha, "Synthesis and Biophysical Characterization of Oligonucleotides Containing a 4'-Selenonucleotide," *Journal of the American Chemical Society* 130 (2008): 8578–9.

[4]L. C. Bock, L. C. Griffin, J. A. Latham, E. H. Vermaas, and J. J. Toole, "Selection of Single-Stranded DNA Molecules That Bind and Inhibit Human Thrombin," *Nature* 355, no. 6360 (1992): 564–6; C. G. Peng and M. J. Damha, "G-Quadruplex Induced Stabilization by 2'-deoxy-2'-fluoro-D-arabino-nucleic Acids (2'F-ANA)," *Nucleic Acids Research* 35, no. 15 (2007): 4977–88.

[5]R. F. Macaya, P. Schultze, F. W. Smith, J. A. Roe, and J. Feigon, "Thrombin-Binding DNA Aptamer Forms a Unimolecular Quadruplex Structure in Solution," *Proceedings of the National Academy of Sciences of the USA* 90, no. 8 (1993): 3745–9.

[6]Because the four bases have different masses, we can calculate the predicted mass of our desired oligonucleotide. If the mass is correct, it does not confirm the sequence of the bases, but because of the nature of the chemical synthesis cycle, nucleotides are not usually delivered in the incorrect order. If our synthesis fails, it is much more likely to produce a truncated sequence (i.e., the oligonucleotide missing one or more nucleotides) and will thus be of incorrect mass.

[7]R. Isaac, "Information, Intelligence, and the Origins of Life," *Perspectives on Science and Christian Faith* 63, no. 4 (2011): 219–30.

[8]W. A. Dembski, *The Design Inference: Eliminating Chance through Small Probabilities* (New York: Cambridge University Press, 1998).

[9]For example, see Meyer, *Signature in the Cell,* chapter 13.

[10]D. R. Venema, "Seeking a Signature," *Perspectives on Science and Christian Faith* 62, no. 4 (2010): 276–83; several responses are available online, including Darrel Falk's review of the book at http://biologos.org/blog/signature-in-the-cell and a multipart series by Dennis Venema beginning at http://biologos.org/blog/evolution-and-origin-of-biological-information-part-1-intelligent-design.

[11]For evidence that natural selection can generate new *genes*, see C. Chandrasekaran and E. Betrán, "Origins of New Genes and Pseudogenes," *Nature Education* 1, no 1 (2008), available online at http://www.nature.com/scitable/topicpage/origins-of-new-genes-and-pseudogenes-835.

Novel *functions* can arise through natural selection, as shown in Lenski's long-term evolution experiment; see N. Philippe, L. Pelosi, R. E. Lenski, and D. Schneider, "Evolution of Penicillin-Binding Protein 2 Concentration and Cell Shape during a Long-Term Experiment with *Escherichia coli*," *Journal of Bacteriology* 191, no. 3 (2008): 909–21; Z. D. Blount, C. Z. Borland, and R. E. Lenski, "Historical Contingency and the Evolution of a Key Innovation in an Experimental Population of *Escherichia coli*," *Proceedings of the National Academy of Sciences of the USA* 105, no. 23 (2008): 7899–906.

For evidence that complex *molecular machines* can evolve, see K. Miller, "The Flagellum Unspun," http://www.millerandlevine.com/km/evol/design2/article.html.

New information arises through *whole-genome duplication* events and subsequent independent evolution of the two copies. See, for example, P. Dehal and J. L. Boore, "Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate," *PLoS Biology* 3, no. 10 (2005): 1700–8.

For a perspective on the evolution of information associated with *developmental programs and body plans*, see Sean Carroll's book *Endless Forms Most Beautiful* (New York: Norton, 2005).

[12]C. Story, "The God of Christianity and the G.O.D. of Immunology," *Perspectives on Science and Christian Faith* 61, no. 4 (2009): 221–32.

[13]While Meyer does discuss biological evolution to some extent in *Signature in the Cell,* he restricts his version of the law of conservation of information as follows: "In a nonbiological context, the amount of specified information initially present in a system, S_i, will generally equal or exceed the specified information content of the final system, S_f." See Meyer, *Signature in the Cell,* chapter 13.

[14]This is sometimes called "information capacity" or Shannon information.

[15]Bock, Griffin, Latham, Vermaas, and Toole, "Selection of Single-Stranded DNA Molecules That Bind and Inhibit Human Thrombin." The oligonucleotide sequences in the actual experiment were much longer—60 nucleotides of randomized sequence surrounded by 18 nt of known sequence on each end (so 96 nt total). The known-sequence ends are required for PCR (polymerase chain reaction) amplification (the implications of this for information input are discussed later in my article). After five rounds of SELEX, the surviving 96-nt oligonucleotides were sequenced, and the researchers observed that a 15-nt motif was responsible for the high binding affinity.

[16]As a lighter example of SELEX in practice, I could mention that Maureen McKeague won *Science* magazine's 2010 "Dance Your PhD" contest by choreographing the *in vitro* selection of a DNA sequence that binds homocysteine. See http://news.sciencemag.org/sciencenow/2010/10/and-the-dance-your-phd-winner-is.html

[17]Gerald F. Joyce, "Forty Years of *in Vitro* Evolution," *Angewandte Chemie International Edition* 46, no. 34 (2007): 6420–36.

[18]A. D. Ellington and J. W. Szostak, "*In Vitro* Selection of RNA Molecules That Bind Specific Ligands," *Nature* 346, no. 6287 (1990): 818–22; G. F. Joyce, "Amplification, Mutation and Selection of Catalytic RNA," *Gene* 82, no. 1 (1989): 83–7; C. Tuerk and L. Gold, "Systematic Evolution of Ligands by

Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase," *Science* 249, no. 4968 (1990): 505–10.

[19]N. K. Navani and Y. F. Li, "Nucleic Acid Aptamers and Enzymes as Sensors," *Current Opinion in Chemical Biology* 10, no. 3 (2006): 272–81.

[20]E. W. M. Ng, D. T. Shima, P. Calias, E. T. Cunningham, D. R. Guyer, and A. P. Adamis, "Pegaptanib, a Targeted Anti-VEGF Aptamer for Ocular Vascular Disease," *Nature Reviews Drug Discovery* 5, no. 2 (2006): 123–32.

[21]See Meyer, *Signature in the Cell,* chapters 13–14 and Appendix A. These references focus specifically on ribozyme engineering, a sub-type of SELEX.

[22]The oligonucleotides are amplified after each step by the polymerase chain reaction (PCR). This famous technique makes use of a polymerase enzyme that requires a primer to start the synthesis of each copy it makes. So the wings of known sequence surrounding our random oligonucleotide are primer binding sites (each one is complementary to a short primer, and primers must be added during the amplification step). In a similar way, DNA sequencing makes use of a polymerase enzyme and thus requires a primer binding site of known sequence.

[23]C. Deck, M. Jauker, and C. Richert, "Efficient Enzyme-Free Copying of All Four Nucleobases Templated by Immobilized RNA," *Nature Chemistry* 3, no. 8 (2011): 603–8; T. A. Lincoln and G. F. Joyce, "Self-Sustained Replication of an RNA Enzyme," *Science* 323, no. 5918 (2009): 1229–32.

[24]K. M. Esvelt, J. C. Carlson, and D. R. Liu, "A System for the Continuous Directed Evolution of Biomolecules," *Nature* 472 (2011): 499–503; M. C. Wright and Gerald F. Joyce, "Continuous *in Vitro* Evolution of Catalytic Function," *Science* 276 (1997): 614–7.

[25]Calculated as the base 2 logarithm of the number of possible states, 4^{20} or 1.1×10^{12}.

[26]The interested reader can explore work from researchers such as Larry Gold (U. Colorado at Boulder), David Liu (Harvard), Gerald Joyce (Scripps), Andy Ellington (U. Texas at Austin), Yingfu Li (McMaster), Frances Arnold (Caltech), and many others, along with various companies whose focus is generating useful nucleic acid structures by *in vitro* evolution and applying them as diagnostic tools, therapeutics, molecular biology reagents and so on: SomaLogic, Aptagen, Archemix, and others. Hundreds of evolved functional sequences are indexed and catalogued by the Ellington lab at http://aptamer.icmb.utexas.edu/.

[27]The concepts in this section are discussed more fully and in their application to biological evolution in the companion article by Stephen Freeland, "The Evolutionary Origins of Genetic Information," *Perspectives on Science and Christian Faith* 63, no. 4 (2011): 240–54.

[28]M. W. Kanan, M. M. Rozenman, K. Sakurai, T. M. Snyder, and D. R. Liu, "Reaction Discovery Enabled by DNA-Templated Synthesis and *in Vitro* Selection," *Nature* 431, no. 7008 (2004): 545–9; M. M. Rozenman, M. W. Kanan, and D. R. Liu, "Development and Initial Application of a Hybridization-Independent, DNA-Encoded Reaction Discovery System Compatible with Organic Solvents," *Journal of the American Chemical Society* 129 (2007): 14933–8.

[29]Ibid.

[30]F. H. Arnold, "Design by Directed Evolution," *Accounts of Chemical Research* 31 (1998): 125–31.

[31]For example, see Meyer, *Signature in the Cell,* chapter 14.

[32]J. S. Wilkins and W. R. Elsberry, "The Advantages of Theft over Toil: The Design Inference and Arguing from Ignorance," *Biology and Philosophy* 16, no. 5 (2001): 709–22.

[33]H. S. Wilf and W. J. Ewens, "There's Plenty of Time for Evolution," *Proceedings of the National Academy of Sciences of the USA* 107, no. 52 (2010): 22454–6.

[34]Meyer, *Signature in the Cell,* chapter 17.

[35]C. Deck, M. Jauker, and C. Richert, "Efficient Enzyme-Free Copying of All Four Nucleobases Templated by Immobilized RNA," *Nature Chemistry* 3, no. 8 (2011): 603–8; T. A. Lincoln and G. F. Joyce, "Self-Sustained Replication of an RNA Enzyme," *Science* 323, no. 5918 (2009): 1229–32; T. R. Cech, "The RNA Worlds in Context," *Cold Spring Harbor Perspectives in Biology* (2011); P. C. Joshi, M. F. Aldersley, J. W. Delano, and J. P. Ferris, "Mechanism of Montmorillonite Catalysis in the Formation of RNA Oligomers," *Journal of the American Chemical Society* 131, no. 37 (2009): 13369–74; M. W. Powner, B. Gerland, and J. D. Sutherland, "Synthesis of Activated Pyrimidine Ribonucleotides in Prebiotically Plausible Conditions," *Nature* 459, no. 7244 (2009): 239–42.

[36]Venema has drawn attention to the particular danger of gap-based arguments in a rapidly evolving field of research: see "Intelligent Design, Abiogenesis, and Learning from History: A Reply to Meyer," *Perspectives on Science and Christian Faith* 63, no. 3 (2011): 183–92.